

Some Theoretical Considerations on Resource-Bounded Defeasible Argumentation¹

Carlos I. Chesñevar²

Guillermo R. Simari

Instituto de Ciencias e Ingeniería de Computación (ICIC)

Grupo de Investigación en Inteligencia Artificial (GIIA)

Departamento de Ciencias de la Computación

Universidad Nacional del Sur

Av. Alem 1253 – (8000) Bahía Blanca – REPÚBLICA ARGENTINA

FAX:(54)(91)563401 – PHONE: (54)(91)20776 (ext.208) – Email:{grs,ccchesne}@criba.edu.ar

KEYWORDS: artificial intelligence, defeasible argumentation, defeasible reasoning

Abstract

During the last decade, defeasible argumentation has proven to be a fruitful formalization for modelling commonsense reasoning. In this respect, the MTDR framework has come to be a fairly standard approach, because of its simplicity and expressive power. Recently, the framework has been enriched through a number of dialectical considerations, resulting in a new ontology, based on Rescher's model of theory formation through dispute.

The current trend is that dispute is the most fair and effective way to investigate the tenability of claims. However, dispute is usually resource-bounded, so that the notion of ultimately supported claims (or justifications) cannot be separated from this issue. This paper discusses some theoretical considerations for an extension of MTDR where bounds on resources are taken into account.

¹This work was partially supported by the Secretaría de Ciencia y Técnica de la Univ. Nac. del Sur

²Supported by a fellowship of the Consejo Nacional de Investigaciones Científicas y Técnicas (CON-ICET), República Argentina.

1 Introduction and motivations

Argumentative systems [SL92, Vre93, SCG94]³ are formalizations of the process of defeasible reasoning. An argument A for a claim h is a tentative piece of reasoning an agent would be inclined to accept, all things considered, as an *explanation* for h .

Arguments can be thought of as defeasible proofs built from the agent's knowledge base. An argument A whose conclusion is h (denoted $\langle A, h \rangle$) is a defeasible proof for the tenability of h . Any argument $\langle B, j \rangle$ is a defeasible proof *against* the tenability of $\langle A, h \rangle$ whenever accepting both $\langle A, h \rangle$ and $\langle B, j \rangle$ would lead to an inconsistency. Such an argument $\langle B, j \rangle$ is said to be a *counterargument* for $\langle A, h \rangle$. Should $\langle B, j \rangle$ be *preferred* over $\langle A, h \rangle$ for some reason (e.g., specificity or because of some extra-logical reasons), then $\langle B, j \rangle$ is called a *defeater* of $\langle A, h \rangle$. Since defeaters are arguments, they are also subject to defeat by other arguments. This allows us to consider a recursive procedure, in which arguments, defeaters, defeaters of defeaters, and so on, should be taken into account. This procedure can be summarized as follows: a) arguments with no defeaters are *acceptable*; b) an argument is acceptable only if it has no acceptable defeaters. A claim h will be finally supported iff there exists an *acceptable argument* supporting h . Such an argument is called a *justification* for h .

The former process can be easily described in *dialectical* terms [Res77]. The agent's KB represents a *shared basis*,⁴ from which two parties, *proponent* and *opponent*, contend on the acceptability of a given claim h . The burden of proof lies initially on the proponent, who must advance an argument A for accepting h . The opponent's role is to *defeat* that argument, by advancing another argument $\langle B, j \rangle$. Both parties perform *moves*, advancing alternatively a new argument which defeats some of the other party's previous arguments. The process ends when no more moves can be performed; the original argument $\langle A, h \rangle$ advanced by the proponent is a justification if it remains undefeated when moves have been exhausted.

In *A Mathematical Treatment of Defeasible Reasoning* [SL92], or *MTDR*, a clear and theoretically sound framework for defeasible argumentation was introduced. That framework became later enriched through a number of dialectical considerations [SCG94], evolving into a refined, *dialectics*-based argumentation system, which has come to be a fairly standard approach because of its simplicity and expressive power. Defeasible argumentation within *MTDR* (as well as in other argumentative systems) equates elegantly a procedural and a declarative approach. Inference is performed by assuming *unbound resources* for computation, since arguments are exhaustible. Hence, it can be shown whether any given claim is to be ultimately accepted (i.e., *justified*) by following an effective procedure.

However, recent research [Vre95, Lou92, SC95] on formal representations of argument and dispute has led to the conclusion that bounds on computational resources play an important role in defeasible reasoning, particularly when defining *protocols* for interaction among arguments. The importance of protocols in argumentative reasoners, as well as the need to formalize the features of those protocols using an expressive notation, has been stressed by Loui in [Loui92]. In his opinion, reorientation in the KR community is needed to include the formal study of protocol as a part of the study of nonmonotonic reasoning.

³We assume that the reader is familiarized with basic issues in defeasible argumentation. The reader is referred to [Vre93] for a in-depth treatment of the subject.

⁴This term was coined by Ronald Loui.

Dialectics refers to a form of disputation in which a *serializable resource* is alternatively shared, so that one party's use of that resource is informed by the result of the other party's prior use of that resource. That resource has typically been time spent in the search of new arguments. Until now, formalizations of *MTDR* have not taken resource bounds into account.

This paper discusses some theoretical issues on modelling resource-bounded defeasible argumentation within the *MTDR* framework. We outline a formal setting using *multilanguage systems*, which results particularly sensible for a proper definition of protocols and interaction among parties in a dialectical debate. The paper is structured as follows. Section 2 presents the basic notions on multilanguage systems, or ML-systems. Section 3 sketches a ML-system for defeasible argumentation, called MLAR. Section 4 presents the main contributions of the paper. Some new definitions are introduced, and a generic algorithm for characterizing the process of debate is described. We discuss also the notion of *resource-bound negation* and the characterization of protocols within the MLAR system. Finally, section 7 presents the most important conclusions that have been obtained.

2 Multilanguage systems (ML-systems)

Multilanguage systems (ML-systems)⁵ allow us to structure knowledge as a set of different theories, each of them characterized by a formal language. ML-systems allow us to formalize deduction within a single theory as well as *among* theories using a new kind of inference rules called *bridge rules*. In these rules, formulas used in the premises and conclusion may belong to different languages. Bridge rules allow to derive a theorem in a theory from a theorem in another theory, thus allowing communication among them. One of the main advantages of ML-systems is that they can be used for defining sensible formalizations of the notion of *context* [McC80]. Facts and rules can be clustered into subtheories, which capture the structure of a given problem, thus allowing an easier and more natural approach to developing knowledge-base systems.⁶

DEFINITION 2.1 *Multilanguage system (ML).* A *multilanguage system* (or MLS) is a 3-uple $\langle \mathcal{L}, \mathcal{A}, \mathcal{R} \rangle$, whose components are defined as follows:

Languages: The set $\mathcal{L} = \{L_i\}_{i \in I}$ is a family of *languages*, where I denotes a set of indices. We will write $L_i:f$ for denoting the fact that f is a wff in L_i .⁷ In that case, we will also say that f is a L_i -wff. A set of L_i -wffs will be denoted S_i .

Axioms: For every $L_i \in \mathcal{L}$ we distinguish a subset of wffs $A_i \subseteq L_i$, called the *axioms* of L_i . The set $\mathcal{A} = \{A_i : A_i \subseteq L_i, i \in I\}$ represents the *axioms* of the system.

Inference rules: $\mathcal{R} = \{R_1, R_2, \dots, R_m\}$ is set of inference rules. Each rule R_i has the form

$$\frac{S_{i_1}, S_{i_2}, \dots, S_{i_m}}{L_j:f}$$

⁵We are aware that the abbreviation 'ML' might lead to misunderstanding, since it is also used for *meta-level*.

⁶The reader is referred to [Giu91, GS94] for further details on multilanguage systems.

⁷It must be noted that our formalizations differs partially from Giunchiglia's original definitions [Giu91]. In particular, he writes $\langle A, i \rangle$ for denoting that A is a wff in the language L_i . We did not use that notation since it might be confused with the one used for argument structures (see appendix).

where $i_k \in I, k = 1 \dots m$. Thus, the *premises* of a given inference rule R_i are given as wffs which may belong to distinct languages. The *conclusion* of R_i is a wff f which belongs to a particular language L_j . \square

DEFINITION 2.2 *Derivability in a MLS.* Let $F = \langle \mathcal{L}, \mathcal{A}, \mathcal{R} \rangle$ be a ML-system. Let S_i be a set of wffs in $L_i \in \mathcal{L}$. Let $\Gamma = \bigcup \{S_j\}, j \in I$, where I is a set of indices associated with F . We will say that the wff $L_i:f$ is *derivable in F from Γ* , written $\Gamma \vdash_F L_i:f$, if there exists a sequence p_1, p_2, \dots, p_k , such that for every p_i it holds that either (a) p_i is an instance of an axiom in F , or (b) $p_i \in \Gamma$, or (c) p_i is a consequence of previous p_j 's by virtue of the application of some inference rule in \mathcal{R} . If $\Gamma = \emptyset$ and $\Gamma \vdash_F L_i:f$, we will say that $L_i:f$ is *provable in F* . \square

3 The MLAR system: fundamentals

Next we will present the main features of MLAR, a ML-system for defeasible argumentation. The reader is referred to [CS96] for a complete description. The main idea is to capture the different aspects of the MTDR framework within a multilanguage system. This involves basically two aspects:

- 1) **Knowledge representation:** the knowledge base of an intelligent agent using the MTDR framework will be represented through the *axioms* of the MLAR system. Axioms will correspond to distinguished subsets of two distinct languages L_K and L_Δ .
- 2) **Inference engine:** the elements needed for performing defeasible inference in MTDR will be described in terms of different languages ($L_\sim, L_{Arbit}, L_{Pro}, L_{Opp}, L_{Debate}$), which allow to capture an introspective debate among two parties (proponent and opponent). ruled by a third party called the *arbiter*.

Subindices in each language L_i denote the intended meaning of formulas in L_i . The languages L_K and L_Δ are used for representing non-defeasible and defeasible knowledge, respectively. Formulas in L_\sim will represent defeasible proofs obtained from L_K and L_Δ . Formulas in L_{Pro} and L_{Opp} will stand for arguments asserted by the proponent and the opponent as the debate proceeds. Formulas in L_{Arbit} will capture the arbiter's knowledge on the debate, establishing when a defeasible proof is an argument, when an argument defeats another, etc. Finally, the different *stages* in the debate [SC95] are captured through formulas in L_{Debate} . Every stage accounts for a number of arguments being undefeated (*alive*), and a number of arguments being defeated (*dead*). Stages are numbered consecutively. We characterize defeasible inference as a *proof* within a ML-system.

DEFINITION 3.1 (*The MLAR system*) A ML-system for defeasible argumentation is defined as a 3-uple $\langle \{L_K, L_\Delta, L_\sim, L_{Debate}, L_{Arbit}, L_{Pro}, L_{Opp}\}, \{K, \Delta\}, \mathcal{R} \rangle$ where $K \subseteq L_K$, $\Delta \subseteq L_\Delta$, such that every formula $e \in \Delta$ is non-ground. The set \mathcal{R} of inference rules characterizes the valid inference steps that can be performed. \square

For space limitations, we do not present the whole set of inference rules in MLAR. As an example, some of them are shown in figure 1.

From the languages and the inference rules specified above, a *debate* in dialectical terms can be conceptualized as a *proof sequence* within a ML-system. More formally, an argument $\langle A, h \rangle$ will be a justification iff there exists a proof sequence s_1, \dots, s_n , where the wff s_n denotes that $\langle A, h \rangle$ is a justification.

1. Definition of argument:

$$\frac{L_{\vdash}:\neg\text{nonminimal}(\text{dproof}(A,h)) \quad L_{\vdash}:\neg\text{inconsistent}(\text{dproof}(A,h))}{L_{\text{Arbit}}:\text{arg}(A,h)}$$

2. Definition of counterargument:

$$\frac{\frac{L_{\text{Arbit}}:\text{arg}(A_1,h_1) \quad L_{\text{Arbit}}:\text{arg}(A_2,h_2)}{L_{\text{Arbit}}:\text{subarg}(\text{arg}(A,h),\text{arg}(A_2,h_2))} \quad L_{\text{Arbit}}:\text{disagree}(\text{arg}(A,h),\text{arg}(A_2,h_2))}{L_{\text{Arbit}}:\text{counterargues}(\text{arg}(A_1,h_1),\text{arg}(A_2,h_2))}$$

3. Parties can advance arguments

$$\frac{L_{\text{Arbit}}:\text{arg}(A,h)}{L_{\text{Pro}}:\text{arg}(A,h)} \quad \frac{L_{\text{Arbit}}:\text{arg}(A,h)}{L_{\text{Opp}}:\text{arg}(A,h)}$$

4. Beginning of debate

$$\frac{L_{\text{Pro}}:\text{arg}(A,h)}{L_{\text{Debate}}:\text{debate}(0,\text{arg}(A,h),\text{Pro},\text{alive})}$$

5. Definition of justification:

$$\frac{L_{\text{Debate}}:\text{debate}(i,\text{arg}(A,h),\text{Party},\text{alive}) \quad L_{\text{Debate}}:\neg\text{debate}(\text{succ}(i),\text{arg}(A,h),\text{Party},\text{dead})}{L_{\text{Arbit}}:\text{justified}(\text{arg}(A,h))}$$

Figure 1: Some of the inference rules used in the MLAR system

DEFINITION 3.2 (*justification*) Let $h \in L_K \cup L_\Delta$, where L_K y L_Δ are the languages for knowledge representation in a ML-system for defeasible argumentation $S = \langle \{L_K, L_\Delta, L_{\vdash}, L_{\text{Debate}}, L_{\text{Arbit}}, L_{\text{Pro}}, L_{\text{Opp}}\}, \{K, \Delta\}, \mathcal{R} \rangle$. Then h will be *justified* iff the formula $\text{justified}(\text{arg}(A,h))$ of L_{Arbit} is provable for some A , i.e., if and only if $\vdash_S L_{\text{Arbit}}:\text{justified}(\text{arg}(A,h))$. \square

The following example shows MTDR's behavior for a sample KB, and how it is captured in MLAR:

EXAMPLE 3.1 Let $K = \{e_1, e_2, c \rightarrow c_2\}$, and let $\Delta = \{c_1 \wedge c_2 \vdash h, e_2 \vdash c, e_1 \vdash c_1, e_2 \wedge c_4 \vdash \neg c_1, e_1 \vdash c_4, e_1 \wedge e_2 \vdash \neg c_4\}$. Then $A_1 = \{e_1 \vdash c_1, e_2 \vdash c, c_1 \wedge c_2 \vdash h\}$ is an argument for h . The argument $\langle A_1, h \rangle$ has an associated defeater $\langle A_2, \neg c_1 \rangle$, $A_2 = \{e_1 \vdash c_4, c_4 \wedge e_2 \vdash \neg c_1\}$. Then $\langle A_1, h \rangle$ is defeated, and cannot (ultimately) be accepted. But this second argument is not acceptable either, since it has a defeater $\langle A_3, \neg c_4 \rangle$, $A_3 = \{e_1 \wedge e_2 \vdash \neg c_4\}$. Hence $\langle A_1, h \rangle$ is reinstated. There are no more arguments to take into consideration. Then $\langle A_3, \neg c_4 \rangle$ is a *justification* (it is undefeated); $\langle A_2, \neg c_1 \rangle$ is not acceptable, because of being defeated (so it is not a justification). The original argument $\langle A_1, h \rangle$ is acceptable (since it has no acceptable defeaters), being also a justification.

In MLAR, the whole process is represented as a proof sequence s_1, s_2, \dots, s_n , being $s_n = L_{\text{Arbit}}:\text{justified}(\text{arg}(A_1,h))$. The provisional acceptance (rejection) of an argument $\langle A, h \rangle$ presented by *Party* (proponent or opponent) would be captured by a wff s_k in that sequence having the form $L_{\text{Debate}}:\text{debate}(i,\text{arg}(A,h),\text{Party},\text{alive})$ ($L_{\text{Debate}}:\text{debate}(i,\text{arg}(A,h),\text{Party},\text{dead})$). The index i denotes the current *stage* of the debate. Provisional acceptance (rejection) may result in rejection (acceptance) in further stages of the debate (see [CS96] for details).

4 Resource-bounded defeasible argumentation

In standard *MTDR*-based approaches, the process of determining whether there exists a justification for a given claim h can be characterized through an *effective* procedure [SL92], which can be modelled as a debate between two parties. Termination is guaranteed, having the same outcome independently on *how* the debate is performed. Inference in MLAR is sound and complete with respect to *MTDR*, i.e. $\langle A, h \rangle$ is a justification in *MTDR* if and only if there exists a proof sequence s_1, s_2, \dots, s_n in MLAR such that $s_n = L_{Arbit} \text{ justified}(\arg(A, h))$.

In [Lou92], Ronald Loui considers the role of *bounds* on computational resources with respect to defeasible argumentation. As he points out, rules in conflict should be regarded as policies, which are inputs to deliberative processes. Dialectical protocols would be appropriate for such deliberations when resources are bounded. One of the main contributions of his paper is the observation that non-monotonic reasoning can be subsumed in a dialectical framework, where the notion of process and non-determinism will play a fundamental role.

Loui discusses a “skeletal model” for characterizing a resource-bounded defeasible reasoner. In that model, a debate is defined as a *locution record*, a finite sequence $\Gamma = \langle l_1, \dots, l_k \rangle$. Each locution is a 3-uple $\langle p_i, s_i, r_i \rangle$, where p_i denotes a *party*, s_i is a *sentence*, and r_i is a “resource marker”. Quoting Loui:

[..] each p_i is a party, from a set P of parties (some of which are players); each s_i is either a wff in a language L (which may actually be better thought to be a set of languages, such as an object language and its metalanguage), or is a structure meeting some well-formedness conditions, such as being an argument for a sentence in L ; and where r_i is a description of resources, such as a d -vector, a marker of resources consumed, of d different resource types.

Loui goes on by defining a number of distinguished elements in that skeletal model. The most important one is the notion of *disputation protocol*, which consists of different functions for making locutions, representing locution-obligations as well as locution-restrictions, and determining when the debate has finished and which is its outcome. This characterization gives a wide-ranging, semi-formal setting for understanding resource-bounded defeasible argumentation and conceptualizing existing argumentative systems inside that setting.

We contend that the MLAR system allows a sensible formalization of Loui’s approach, without losing *MTDR*’s sound theoretical background. A *proof* in MLAR can be namely seen as a locution record; every locution $\langle p_i, s_i, r_i \rangle$ can be characterized as a sequence of wffs in that proof. The notion of context, formalized in terms of a ML-system, provides a generic framework for defining flexible interaction among inference rules. We want to extend *MTDR*, computing justifications as the outcome of a resource-bounded process.

A debate can be thought of as a *resource-bounded process* for determining the acceptability of a claim h , whose outcome is a *set* of logical propositions. That set would be called *current opinion*.⁸ The basic algorithm for a debate process could resemble the one shown in figure 2. First, a party p_i must be chosen to perform a move m_i . That move

⁸This term was introduced in Loui’s skeletal model [Lou92], although in a different sense. We use this term following Vreeswijk [Vre95].

will consume r_i computational resources from party p_i , and change the current opinion.⁹ The debate would be carried on until either the original claim h is *established* (justified)¹⁰ or resources are exhausted. In the latter case, both expenditure of resources and current opinion would be a measure for the “justification degree” of the claim. Next we will introduce some definitions, which can be helpful by providing a more formal insight:

Algorithm: Debate_process
 INPUT: Claim h
 OUTPUT: *Current_opinion*
Repeat
 Select *party* p_i to perform next *move* m_i .
 Performing move m_i consumes r_i *resources*,
 and changes *current opinion*.
Until (h is *established*) **or** (resources exhausted)

Figure 2: The basic algorithm for performing a resource-bounded debate

DEFINITION 4.1 (*debate*) A *debate* d is a finite sequence of *moves*, $d = \{m_1, m_2, \dots, m_k\}$. A debate is *exhaustive* if no further moves are allowed after move m_k has been carried out. \square

Clearly, in standard *MTDR* all exhaustive debates lead to the same outcome (even though some of them can be preferred over others, as we will see next). However, when resources are bounded, exhaustive debates may differ, since expenditure of resources is also an input to the deliberative process. Next we will state Loui’s notion of *move* within MLAR:

DEFINITION 4.2 (*move*) Let $S = \langle \mathcal{L}, \mathcal{A}, \mathcal{R} \rangle$ be a MLAR system. A *move* m_i is a 3-uple $\langle p_i, s_i, r_i \rangle$, where

- p_i denotes a *party*. There are three possible parties which may perform moves: *proponent*, *opponent*, and *arbiter*.
- f_i denotes a wff that can be inferred from wffs in previous moves m_1, \dots, m_{i-1} using the inference rules in \mathcal{R} , i.e. $\{f_1, \dots, f_{i-1}\} \vdash_s f_i$.
- r_i denotes the computational resources consumed by p_i for inferring f_i .

\square

EXAMPLE 4.1 Consider example 3.1. A first move m_1 could result in $\langle \text{Pro}, L_{\text{Debate}}: \text{debate}(1, \text{arg}(A_1, h), \text{Pro}, \text{alive}), n_1 \rangle$ (n_1 denotes resource units spent in performing the move).¹¹ A second move m_2 could on its turn result in $\langle \text{Pro}, L_{\text{Debate}}: \text{debate}(2, \text{arg}(A_2, \neg c_1), \text{Opp}, \text{alive}), n_2 \rangle$. The arbiter could have the burden of acting after proponent and opponent have advanced arguments. In that case, its move could be determining that the second argument defeats the first, so that the first argument is reinstated as *dead*: $\langle \text{Arbit}, L_{\text{Debate}}: \text{debate}(3, \text{arg}(A_1, h), \text{Pro}, \text{dead}), n_3 \rangle$.

⁹ Assuming of course that there are at least r_i resource units available for p_i

¹⁰ Following the literature, the terms *established*, *justified*, and *warranted* can be used interchangeably, unless stated otherwise.

¹¹ We can assume that all moves have uniform cost. Resource units can be conceptualized as ‘weights’ assigned to inference rules.

Current opinion denotes the set of all logical propositions in L_{Debate} derived so far, representing the current outcome of the debate process. In the example 4.1, current opinion is that $\langle A_2, \neg c_1 \rangle$ prevails, whereas $\langle A_1, h \rangle$ is rejected.

DEFINITION 4.3 (*current opinion*). Let $d = \{m_1, m_2, \dots, m_k\}$ be a debate, where every move m_i has the form $\langle p_i, f_i, r_i \rangle$, as defined above. The *current opinion* of the debate d , denoted $CO(d)$, is the set $\{f_1, f_2, \dots, f_k\}$. \square

Established (or *justified*) opinion denotes the set of all logical propositions in L_{Debate} derived so far, such that they will prevail up to the end of the debate process. Established wffs account for arguments that cannot be defeated in further stages of the debate, *i.e.*, they are justifications.

DEFINITION 4.4 (*established opinion*). Let $d = \{m_1, m_2, \dots, m_k\}$ be a debate, and let $CO(d)$ be its associated current opinion. The *established opinion* of the debate d , denoted $EO(d)$, is the set of all *established* wffs in d . A wff $f_i \in CO(d)$ (derived in move m_i) is said to be *established* iff $\neg f_i$ (derived in move m_j , $j > i$) does not belong to $CO(d)$. \square

In the example 3.1, $\langle A_2, \neg c_1 \rangle$ is not established, since carrying on the debate process it will result defeated.

In standard *MTDR*, given a knowledge base (\mathcal{K}, Δ) , all exhaustive debates that can be performed have the *same* length, and the established opinion will be always the same. Can we have some kind of preference among different exhaustive debates? Determining as soon as possible whether the original claim is justified might provide a preference criterion:

DEFINITION 4.5 (*Preference among debates*). Let d_i, d_j be two exhaustive debates about a given claim h . The debate d_i is said to be *preferred over* d_j iff 1) There exists an argument $\langle A, h \rangle$ such that the wff $\text{justified}(A, h)$ can be established in debate d_i in k_i moves, and 2) For every argument $\langle B, h \rangle$, such that the wff $\text{justified}(B, h)$ can be established in debate d_j in k_j moves, it holds that $k_j \geq k_i$. \square

Bounds on resources impose move restrictions. Some protocols may be acceptable to reach the “right” outcome, and others may not. The point is whether we can define some desiderata to be satisfied by protocols with respect to the final outcome [Lou92]. This prompts the following question: *Can the “correctness” of the outcome be ensured by the “correctness” of the protocol?*

As we have seen, even though termination can be guaranteed, there still exists preference among exhaustive debates (that is why improving the search strategy in *MTDR* makes sense). When resources are bound, improving the search strategy is essential for good argumentation. This can be observed in colloquial argument [Vre95] carried out in deliberative organs: there are time constraints for speaking, and it is possible to raise “points of order” for steering the debate.

5 Resource-bound negation

Negation has long deserved a deep analysis in knowledge representation and non-monotonic reasoning. Semidecidability of first-order languages led to define the interrelated notions of *closed-world assumption* (CWA) and *negation as failure* (NAF). The latter turned out to be specially relevant within the logic programming community. According to R. Loui [Lou92], NAF can be embedded in a resource-bounded setting. As an example, he considers the case of the two conflicting default rules [Rei80], assuming “time for proof” as the only resource:

$$\frac{A : B}{B} \quad \frac{A : \neg B}{\neg B}$$

Given A , by applying the left rule we can show B , since we are not able to proof $\neg B$ (i.e., it is consistent to assume B). By trying to prove $\neg B$ and failing, after having consumed all resources on the attempted proof. The opposing rule would not be considered. Loui refines then the former rules as *defeasible* rules, which would allow a more sensible treatment of resources in a debate between *proponent* and *opponent*.

- (1) A is a reason for B (2) A is a reason for $\neg B$

Given A , the proponent advances (1) as an argument for B ; if time remains, the opponent tries to resolve the dispute in its favor, by advancing (2). Loui observes that “[...] the NAF example illustrates how NMR, in its early forms, ignores dialectical ideas, resource distribution, and the consideration of fairness under resource-bounded construction that lead to dialectical ideas.”. Finally, he adds “The fairer the protocol of a disputation, and the better the strategic play, and the more effective the expenditure of resource, the better warranted the outcome.”

Traditional logic programming has also been enriched by relating together negation-as-failure and classic negation. This results in so-called *extended* programs [GL90], based on the method of stable models. Extended programs include clauses of the form $\neg Q \leftarrow \text{not} P$, which stands for “if there is no evidence that P is true, then Q is false”. Extended programs are connected with the problem of relation between logic programming and nonmonotonic formalism. As the authors point out in [GL90], “[...] We can say that the class of extended programs is the place where logic programming meets default logic halfway.”

We believe that an MLS-based formalism as the one presented in this paper allows us to nicely combine Loui’s ideas with negation in extended logic programs, resulting in a sort of “resource-bounded negation”. Consider, for example, the following inference rule of the MLAR system:

$$\frac{L_{\text{Debate}}:\text{debate}(i, \arg(A, h), \text{Party}, \text{alive}) \quad L_{\text{Debate}}:\neg\text{debate}(\text{succ}(i), \arg(A, h), \text{Party}, \text{dead})}{L_{\text{Arbit}}:\text{justified}(\arg(A, h))}$$

This rule states that we can conclude that the argument $\langle A, h \rangle$ is a justification only if $\langle A, h \rangle$ has been established by *Party* (proponent) at stage i of the debate, and it is not possible to show that $\langle A, h \rangle$ would be defeated at stage $i+1$. The last statement should be understood under CWA, since there exists an effective procedure for demonstrating whether $\text{debate}(\text{succ}(i), \arg(A, h), \text{Party}, \text{dead})$ holds.

However, proving this sentence favors the opponent’s position; it should not be in charge of the proponent to spend resources (time, memory, etc.) to prove it. We could therefore distinguish two kinds of resources: a) proponent’s time for proof, and b) opponent’s time for proof. Whenever the justification process is carried out, proving a sentence P should be considered as expenditure of a proponent’s (opponent’s) time whenever P supports (undermines) the proponent’s (opponent’s) position. In that case, proving $\text{debate}(\text{succ}(i), \arg(A, h), \text{Party}, \text{dead})$ should be in charge of the opponent.¹²

¹²The relation between extended logic programs and defeasible argumentation was already introduced in [SG95]. However, that approach has a different motivation; assignment of resources is not considered.

6 Protocols and resource bounds

A *protocol* is a kind of 2-party procedure which encodes dialectical information [Vre95]. When computing justifications, many alternative protocols might be followed. Some protocols allow parties to construct many arguments at a time; others don't; some protocols set the burden of the proof always on the proponent, others on both proponent and opponent; and so on. The study of protocols in defeasible argumentation is relatively new; relevant work on this subject can be found in [Vre95, Lou92].

Fairness and effectiveness are basic properties demanded on protocols. Fairness demands maximal opportunity for response: a protocol is *fair* whenever any claim established in dispute is actually justified. Effectiveness demands maximal information about the goal to reach when performing inference: a protocol is *effective* whenever any justified claim can be established in dispute (see [Vre95, Lou92] for a detailed treatment of the subject).

Protocols can be defined *within* an MLAR system through inference rules which specify how parties interact along the course of the debate. For example, for an alternating order *Pro-Opp-Arbit*, as suggested in section 4, the following three rules would suffice:

$$\frac{L_{Arbit}:next_player(arbit)}{L_{Arbit}:next_player(pro)} \quad \frac{L_{Arbit}:next_player(pro)}{L_{Arbit}:next_player(opp)} \quad \frac{L_{Arbit}:next_player(opp)}{L_{Arbit}:next_player(arbit)}$$

It can be the case that expenditure of resources or current opinion (*i.e.*, the proof sequence derived so far) affect the protocol itself. Consider for example the case that we want to allow proponent to attack twice whenever a certain situation γ holds (that situation can be captured through a wff in MLAR). In order to do this, an inference rule as the following could be introduced:

$$\frac{L_{Arbit}:next_player(pro) \wedge \gamma}{L_{Arbit}:next_player(pro)}$$

7 Conclusions and future work

In this paper, we have discussed several issues concerning resource-bounded defeasible argumentation, many of which represent problems still open in computational dialectics. We have also shown that a multilanguage system for defeasible argumentation such as MLAR can be properly extended for dealing with theoretical issues concerning bounds on resources.

It is clear that there exists a tradeoff between desirable mathematical properties of [SL92] (such as the existence of an effective procedure for computing justifications) and a non-demonstrative, resource-bounded approach (which might be more adequate for solving real-world problems through defeasible argumentation). Figure 3 sketches some of the main distinctive features between standard and resource-bounded *MTDR*.

It is still debatable to what extent we should include non-determinism within an argumentative system as an input that affects the final outcome of a deliberative process. It should be observed that recent research [Che94] has demonstrated that protocol issues are relevant even when considering no bounds on resources, leading to important computational savings. Notions such as commitment store and burden of proof deserve as

- | | |
|---|---|
| <ul style="list-style-type: none"> • Any protocol is both fair and effective. • All exhaustive debates are equal. • The current outcome of the process varies; the final outcome is always the same. | <ul style="list-style-type: none"> • Effectiveness and fairness depend on protocol definitions. • Exhaustive debates may be different. • Both current and final outcome of the process may vary. |
|---|---|

Figure 3: Standard and resource-bounded *MTDR*: a comparison

well a more formal treatment in order to fully incorporate them in existing argumentative systems.

Even though computational dialectics is a relatively new field, much effort has been already spent in studying complex issues such as self-modifying protocols, rule formation, and preference criteria among arguments [Vre95], sometimes semi-formally. In our opinion, a unified formal setting as the one outlined in this paper provides the necessary theoretical background for studying several properties and features desirable in actual argumentative systems.

A The *MTDR* framework

We will briefly introduce the main concepts and definitions of the *MTDR* framework (see [SL92, SCG94] for further details).

Knowledge representation: The knowledge of an intelligent agent \mathcal{A} will be represented using a first-order language \mathcal{L} , plus a binary meta-linguistic relation “ \succ ” between sets of non-ground literals of \mathcal{L} which share variables. The members of this meta-linguistic relation will be called *defeasible rules*, and they have the form “ $\alpha \succ \beta$ ”. The relation “ \succ ” is understood as expressing that “reasons to believe in the antecedent α provide reasons to believe in the consequent β ”. Let \mathcal{K} be a consistent subset of sentences of the language \mathcal{L} . This set can be partitioned in two subsets \mathcal{K}_G , of *general* (necessary) knowledge, and \mathcal{K}_P , of *particular* (contingent) knowledge. The beliefs of \mathcal{A} are represented by a pair (\mathcal{K}, Δ) , called *Defeasible Logic Structure*, where Δ is a finite set of defeasible rules. \mathcal{K} represents the non-defeasible part of \mathcal{A} ’s knowledge and Δ represents information that \mathcal{A} is prepared to take at less than face value. Δ^\dagger denotes the set of all ground instances of members of Δ .

Inference

DEFINITION A.1 Let Γ be a subset of $\mathcal{K} \cup \Delta^\dagger$. A ground literal h is a *defeasible consequence* of Γ , abbreviated $\Gamma \vdash h$, if and only if there exists a finite sequence B_1, \dots, B_n such that $B_n = h$ and for $1 \leq i < n$, either $B_i \in \Gamma$, or B_i is a direct consequence of the preceding elements in the sequence by virtue of the application of any inference rule of the first-order theory associated with the language \mathcal{L} . The ground instances of the defeasible rules are regarded as material implications for the application of inference rules. We will also write $\mathcal{K} \cup A \vdash h$ distinguishing the set A of defeasible rules used in the derivation from the context \mathcal{K} . \square

DEFINITION A.2 Given a context \mathcal{K} , a set Δ of defeasible rules, and a ground literal h in the language \mathcal{L} , we say that a subset A of Δ^\dagger is an *argument structure* for h in the context \mathcal{K} (denoted by $\langle A, h \rangle_{\mathcal{K}}$, or just $\langle A, h \rangle$) if and only if: 1) $\mathcal{K} \cup A \vdash h$, 2) $\mathcal{K} \cup A \not\vdash \perp$ and 3) $\nexists A' \subset A, \mathcal{K} \cup A' \vdash h$. Given an argument structure $\langle A, h \rangle$, we also say that A is an argument for h . A *subargument* of $\langle A, h \rangle$ is an argument $\langle S, j \rangle$ such that $S \subseteq A$. \square

DEFINITION A.3 Two argument structures $\langle A_1, h_1 \rangle$ and $\langle A_2, h_2 \rangle$ *disagree*, denoted $\langle A_1, h_1 \rangle \bowtie \langle A_2, h_2 \rangle$, if and only if $\mathcal{K} \cup \{h_1, h_2\} \vdash \perp$. \square

DEFINITION A.4 Given two arguments $\langle A_1, h_1 \rangle$ and $\langle A_2, h_2 \rangle$, we say that $\langle A_1, h_1 \rangle$ *counterargues* $\langle A_2, h_2 \rangle$, denoted $\langle A_1, h_1 \rangle \otimes^h \langle A_2, h_2 \rangle$ iff 1) There exists a subargument $\langle A, h \rangle$ of $\langle A_2, h_2 \rangle$ such that $\langle A_1, h_1 \rangle \bowtie \langle A, h \rangle$; 2) For every proper subargument $\langle S, j \rangle$ of $\langle A_1, h_1 \rangle$, it is not the case that $\langle A_2, h_2 \rangle \otimes \langle S, j \rangle$. \square

DEFINITION A.5 Given two argument structures $\langle A_1, h_1 \rangle$ and $\langle A_2, h_2 \rangle$, we say that $\langle A_1, h_1 \rangle$ *defeats* $\langle A_2, h_2 \rangle$ at literal h , denoted $\langle A_1, h_1 \rangle \gg_{\text{def}} \langle A_2, h_2 \rangle$, if and only if there exists a subargument $\langle A, h \rangle$ of $\langle A_2, h_2 \rangle$ such that: $\langle A_1, h_1 \rangle$ counterargues $\langle A_2, h_2 \rangle$ at the literal h and 1) $\langle A_1, h_1 \rangle$ is *strictly more specific*¹³ than $\langle A, h \rangle$, or 2) $\langle A_1, h_1 \rangle$ is unrelated by specificity to $\langle A, h \rangle$. If $\langle A_1, h_1 \rangle \gg_{\text{def}} \langle A_2, h_2 \rangle$, we will also say that $\langle A_1, h_1 \rangle$ is a *defeater* for $\langle A_2, h_2 \rangle$. \square

We will accept an argument A as a defeasible reason for a conclusion h if A is a *justification* for h . The acceptance of the original argument A as a justification for h will result from a recursive procedure, in which arguments, counterarguments, counter-counterarguments, and so on, should be taken into account. This leads to a tree structure, called *dialectical tree*. Paths along that tree will be called *argumentation lines*, which can be thought of alternate sequences of *supporting* and *interfering* arguments in a debate.

DEFINITION A.6 Let $\langle A, h \rangle$ be an argument structure. A *dialectical tree* for $\langle A, h \rangle$, denoted $\mathcal{T}_{\langle A, h \rangle}$, is recursively defined as follows:

1. A single node containing an argument structure $\langle A, h \rangle$ with no defeaters is by itself a dialectical tree for $\langle A, h \rangle$.
2. Suppose that $\langle A, h \rangle$ is an argument structure with defeaters $\langle A_1, h_1 \rangle, \langle A_2, h_2 \rangle, \dots, \langle A_n, h_n \rangle$. We construct the dialectical tree for $\langle A, h \rangle$, $\mathcal{T}_{\langle A, h \rangle}$, by putting $\langle A, h \rangle$ in the root node of it and by making this node the parent node of the roots of the acceptable dialectical trees of $\langle A_1, h_1 \rangle, \langle A_2, h_2 \rangle, \dots, \langle A_n, h_n \rangle$.

\square

DEFINITION A.7 Let $\mathcal{T}_{\langle A, h \rangle}$ be a dialectical tree for $\langle A, h \rangle$. Nodes in $\mathcal{T}_{\langle A, h \rangle}$ can be recursively labeled as *undefeated nodes* (U-nodes) and *defeated nodes* (D-nodes) as follows: a) Leaves in $\mathcal{T}_{\langle A, h \rangle}$ are U-nodes; b) Let $\langle B, q \rangle$ be an inner node in $\mathcal{T}_{\langle A, h \rangle}$. Then $\langle B, q \rangle$ will be a U-node iff every child node of $\langle B, q \rangle$ is a D-node. $\langle B, q \rangle$ will be a D-node iff it has at least one U-node as a child node. \square

DEFINITION A.8 Let $\langle A, h \rangle$ be an argument structure, and let $\mathcal{T}_{\langle A, h \rangle}$ be its associated dialectical tree.¹⁴ We will say that A is a *justification* for h (or $\langle A, h \rangle$ is a *justification*) iff the root node of $\mathcal{T}_{\langle A, h \rangle}$ is a U-node. \square

¹³Specificity imposes a partial order on argument structures, being used as a preference criterion among them [SCG94]. However, other preference criteria could also be valid.

¹⁴Actually, dialectical trees should satisfy a number of constraints for being considered *acceptable* dialectical trees (see [SCG94]). That issue, however, exceeds the scope of this paper.

References

- [Che94] Carlos I. Chesñevar. Stratifying defeasible rules in argumentative reasoning: a consistency-based approach. En *Primer Workshop Argentino sobre Aspectos Teóricos en Inteligencia Artificial (Buenos Aires)*, Abril 1994.
- [CS96] Carlos Iván Chesñevar y Guillermo R. Simari. *Un sistema multilenguaje para argumentación rebatible*. Reporte Técnico, Grupo de Investigación en Inteligencia Artificial (GIIA), Universidad Nacional del Sur, Bahía Blanca, Argentina. (Submitted to the Segundo Congreso Argentino de Ciencias de la Computación), San Luis, Argentina, Noviembre 1996.
- [Giu91] Fausto Giunchiglia. Multilanguage systems. En *Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense reasoning*, 1991.
- [GL90] Michael Gelfond y Vladimir Lifschitz. Logic programs with classical negation. En *Proceedings of the 7th International Conference on Logic Programming. Jerusalem.*, Junio 1990.
- [GS94] Fausto Giunchiglia y Luciano Serafini. Multilanguage hierarchical logics (or: how we can do without modal logics). *Artificial Intelligence*, (65):29–70, 1994.
- [Lou92] Ronald P. Loui. Process and policy: resource-bounded, non-demonstrative argument. *Computational Intelligence (to appear)*, Octubre 1992.
- [McC80] John McCarthy. Circumscription – A Form of Non-monotonic Reasoning. *Artificial Intelligence*, 13:27–39,171–172, 1980.
- [Rei80] Raymond Reiter. A Logic for Default Reasoning. *Artificial Intelligence*, 13(1,2):81–132, Apr 1980.
- [Res77] Nicholas Rescher. *Dialectics, a Controversy-Oriented Approach to the Theory of Knowledge*. State University of New York Press, Albany, USA, 1977.
- [SC95] Guillermo R. Simari y Carlos I. Chesñevar. Arguments and contexts. En *XXI Conferencia Latinoamericana de Informática*, CLEI/PANEL 95, Canela (Brasil)., Julio 1995.
- [SCG94] Guillermo R. Simari, Carlos I. Chesñevar, y Alejandro J. García. The role of dialectics in defeasible argumentation. En *Anales de la XIV Conferencia Internacional de la Sociedad Chilena para Ciencias de la Computación*, Universidad de Concepción, Concepción (Chile), Noviembre 1994.
- [SG95] Guillermo R. Simari y Alejandro J. García. A knowledge representation language for defeasible argumentation. En *XXI Conferencia Latinoamericana de Informática*, CLEI/PANEL 95, Canela (Brasil)., Julio 1995.
- [SL92] Guillermo R. Simari y Ronald P. Loui. A Mathematical Treatment of Defeasible Reasoning and its Implementation. *Artificial Intelligence*, 53:125–157, 1992.
- [Vre93] Gerard A.W. Vreeswijk. *Studies in Defeasible Argumentation*. Tesis Ph.D., Vrije University, Amsterdam (Holanda), 1993.
- [Vre95] Gerard A.W. Vreeswijk. *Representation of Formal Dispute with a Standing Order*. Reporte Técnico, Department of Computer Science, Rijksuniversiteit Limburg, Maastricht, The Netherlands., 1995.